Sphere Packing Bound for Quantum Channels

Marco Dalai
Department of Information Engineering
University of Brescia, Italy
Email: marco.dalai@ing.unibs.it

***Abstract*—In this paper, the Sphere-Packing-Bound of Fano, Shannon, Gallager and Berlekamp is extended to general classical-quantum channels. The obtained upper bound for the reliability function, for the case of pure-state channels, coincides at high rates with a lower bound derived by Burnashev and Holevo [1]. Thus, for pure state channels, the reliability function at high rates is now exactly determined. For the general case, the obtained upper bound expression at high rates was conjectured to represent also a lower bound to the reliability function, but a complete proof has not been obtained yet.**

I. Introduction

This paper considers the problem of classical communication over quantum channels, focusing on the study of error exponents for optimal codes at rates below the channel capacity. Upper bounds to the probability of error of optimal codes for pure-state channels were obtained by Burnashev and Holevo [1] that are the equivalent of the so-called random coding bound obtained by Fano [2] and Gallager [3] and of the expurgated bound of Gallager [3] for classical channels. The expurgated bound was then extended to general quantum channels by Holevo [4]. The formal extension of the random coding bound expression to mixed states is conjectured to represent an upper bound for the general case but no proof has been obtained yet (see [1], [4]).

In this paper, a sphere packing bound for classical-quantum channels is derived. The quantum case is related to the classical one by means of the *Nussbaum-Szkoła mapping*, introduced in [5] and central to the proof of the converse part of the quantum Chernoff bound (see [6] for more details). This allows us to extend to the quantum case the Shannon-Gallager-Berlekamp generalization of the Chernoff bound introduced in [7] (in its converse part). Then, the proof of the sphere packing bound used in [7] is adapted to the quantum case. This demonstrates the power of the method developed in [7]. Due to space limitation, this paper only includes the main derivation of the results; technical details and additional comments can be found in an extended version of this paper [8].

II. Binary Hypothesis Testing

In this section, the converse part of the Shannon-Gallager-Berlekamp bound for classical binary hypothesis testing [7, Th. 5] is extended to the case of quantum state discrimination. This result will then be used in the next section to derive the sphere packing bound.

Let ϱ and ς be two density operators in a Hilbert space $\mathcal{H}$ and consider the problem of binary hypothesis testing between ϱ and ς. We suppose here that the two density operators have non-disjoint supports, for otherwise the problem is trivial. The decision has to be taken based on the result of a measurement that can be identified with a pair of positive operators $\{\mathbb{1} - \Pi, \Pi\}$, where $0 < \Pi < \mathbb{1}$, associated respectively to ϱ and ς. The probability of error given that the system is in state ϱ or ς is respectively

$$P_{e|\varrho} = \operatorname{Tr} \Pi \varrho \quad \text{and} \quad P_{e|\varsigma} = \operatorname{Tr}(\mathbb{1} - \Pi)\varsigma. \tag{1}$$

Remark 1: This choice of notation is motivated by the fact that our states ϱ and ς do *not* play the role of the states that are usually indicated with ρ and σ in the literature. For example, when comparing Theorem 1 below with the results in [6], we should interpret our quantities with the correspondences $\varrho = \rho^{\otimes N}$ and $\varsigma = \sigma^{\otimes N}$ in mind. Here, however, we will apply the theorem to more general cases where ϱ and ς are tensor products of N not necessarily identical states and, in this sense, Theorem 1 is more general than the results in [6].

Following [7, Sec. 3], for any real s in the interval $0 < s < 1$, define the quantity

$$\mu(s) = \log \operatorname{Tr} \varrho^{1-s} \varsigma^s \tag{2}$$

and let then by definition

$$\mu(0) = \lim_{s \to 0} \mu(s) \quad \text{and} \quad \mu(1) = \lim_{s \to 1} \mu(s). \tag{3}$$

Theorem 1 (Quantum Shannon-Gallager-Berlekamp Bound): Let ϱ, ς be density operators with non-disjoint supports, let Π be a measurement operator for the binary hypothesis test between ϱ and ς, let the probabilities of error $P_{e|\varrho}, P_{e|\varsigma}$ be defined as in (1) and $\mu(s)$ be defined as in (2)-(3). Then, for any $0 < s < 1$, either

$$P_{e|\varrho} > \frac{1}{8} \exp\left[\mu(s) - s\mu'(s) - s\sqrt{2\mu''(s)}\right] \tag{4}$$

or

$$P_{e|\varsigma} > \frac{1}{8} \exp\left[\mu(s) + (1-s)\mu'(s) - (1-s)\sqrt{2\mu''(s)}\right]. \tag{5}$$

Proof: This theorem is essentially the combination of the main idea introduced in [5] for proving the converse part of the quantum Chernoff bound and of [7, Th. 5], the classical version of this same theorem. Since some intermediate steps of those proofs are needed, we unroll the details here for the reader's convenience.

We proceed as in [6]. Let the spectral decomposition of ϱ and ς be respectively

$$\varrho = \sum_i \alpha_i |x_i\rangle\langle x_i| \quad \text{and} \quad \varsigma = \sum_j \beta_j |y_j\rangle\langle y_j|. \tag{6}$$

where $\{|x_i\rangle\}$ and $\{|y_j\rangle\}$ are orthonormal bases. First observe that, from the Quantum Neyman-Pearson Lemma ([9], [10]), it suffices to consider orthogonal projectors Π. So, one has $\Pi = \Pi^2 = \Pi\mathbb{1}\Pi = \sum_j \Pi|y_j\rangle\langle y_j|\Pi$. Symmetrically, we have that $(\mathbb{1}-\Pi) = \sum_i (\mathbb{1}-\Pi)|x_i\rangle\langle x_i|(\mathbb{1}-\Pi)$. Hence, one has

$$\begin{aligned} P_{e|\varrho} &= \operatorname{Tr}\Pi\varrho & (7)\\ &= \sum_{i,j}\alpha_i|\langle x_i|\Pi|y_j\rangle|^2 & (8)\\ P_{e|\varsigma} &= \operatorname{Tr}(\mathbb{1}-\Pi)\varsigma & (9)\\ &= \sum_{i,j}\beta_j|\langle x_i|\mathbb{1}-\Pi|y_j\rangle|^2 & (10)\end{aligned}$$

Using the fact that $|a|^2+|b|^2 \geq |a+b|^2/2$ for any two complex numbers a, b, we find that for all (i,j)

$$\eta_1\alpha_i|\langle x_i|\Pi|y_j\rangle|^2 + \eta_2\beta_j|\langle x_i|\mathbb{1}-\Pi|y_j\rangle|^2 \geq \min(\eta_1\alpha_i,\eta_2\beta_j)\frac{|\langle x_i|y_j\rangle|^2}{2}, \quad (11)$$

which implies that

$$\eta_1 P_{e|\varrho} + \eta_2 P_{e|\varsigma} \geq \frac{1}{2}\sum_{i,j}\min\left(\eta_1\alpha_i|\langle x_i|y_j\rangle|^2, \eta_2\beta_j|\langle x_i|y_j\rangle|^2\right). \quad (12)$$

Now, following [5], consider two probability distributions defined by the Nussbaum-Szkoła mapping

$$P_1(i,j) = \alpha_i|\langle x_i|y_j\rangle|^2, \qquad P_2(i,j) = \beta_j|\langle x_i|y_j\rangle|^2. \quad (13)$$

These two probability distributions are both strictly positive for at least one pair of (i,j) values, since we assumed ϱ, ς to have non-disjoint supports. Furthermore, they have the nice property of allowing for $\mu(s)$, as defined in (2), the expression

$$\mu(s) = \log\sum_{i,j} P_1(i,j)^{1-s}P_2(i,j)^s. \quad (14)$$

Following [7, Th. 5], define the distribution Q_s by

$$Q_s(i,j) = \frac{P_1(i,j)^{1-s}P_2(i,j)^s}{\sum_{i',j'}P_1(i',j')^{1-s}P_2(i',j')^s} \quad (15)$$

and observe that

$$\begin{aligned}\mu'(s) &= E_{Q_s}\left[\log(P_2/P_1)\right] & (16)\\ \mu''(s) &= \operatorname{Var}_{Q_s}\left[\log(P_2/P_1)\right], & (17)\end{aligned}$$

where the subscript Q_s means that the expected values are with respect to the probability distribution Q_s. Hence, if one defines the set

$$Y_s = \left\{(i,j) : \left|\log\left(\frac{P_2(i,j)}{P_1(i,j)}\right) - \mu'(s)\right| \leq \sqrt{2\mu''(s)}\right\} \quad (18)$$

then $\sum_{Y_s} Q_s(i,j) > 1/2$, by Chebyshev's inequality. It is easily checked, using the definitions (15) and (18), that for each $(i,j) \in Y_s$ the distribution Q_s satisfies

$$Q_s(i,j) \leq P_1(i,j)\left(\exp\left[\mu(s) - s\mu'(s) - s\sqrt{2\mu''(s)}\right]\right)^{-1} \quad (19)$$

$$Q_s(i,j) \leq P_2(i,j)\Big(\exp[\mu(s) + (1-s)\mu'(s) - (1-s)\sqrt{2\mu''(s)}]\Big)^{-1}. \quad (20)$$

Hence, in Y_s, $Q_s(i,j)$ is bounded by the minimum of the two expressions on the right hand side of (19) and (20). If we call η_1 the coefficient of $P_1(i,j)$ in (19) and η_2 the coefficient of $P_2(i,j)$ in (20), then we obtain

$$\begin{aligned}\frac{1}{2} &< \sum_{(i,j)\in Y_s} Q_s(i,j) & (21)\\ &\leq \sum_{(i,j)\in Y_s}\min\left(\eta_1 P_1(i,j), \eta_2 P_2(i,j)\right) & (22)\\ &\leq \sum_{(i,j)}\min\left(\eta_1 P_1(i,j), \eta_2 P_2(i,j)\right). & (23)\end{aligned}$$

Now note that the last expression, by the definition of P_1 and P_2 in (13), exactly equals the sum in (12). So, with the selected values of η_1 and η_2 we have $\eta_1 P_{e|\varrho} + \eta_2 P_{e|\varsigma} > 1/4$. But, obviously, $\eta_1 P_{e|\varrho} + \eta_2 P_{e|\varsigma} \leq 2\max\{\eta_1 P_{e|\varrho}, \eta_2 P_{e|\varsigma}\}$. Hence, either $P_{e|\varrho} > \eta_1^{-1}/8$ or $P_{e|\varsigma} > \eta_2^{-1}/8$, concluding the proof. ∎

III. Sphere Packing Bound

Following [4], consider a classical-quantum channel with an input alphabet of K symbols $\{1,\ldots,K\}$ with associated density operators S_k, $k = 1,\ldots,K$ in a finite dimensional Hilbert space $\mathcal{H}$. The N-fold product channel acts in the tensor product space $\mathcal{H}^{\otimes N}$ of N copies of $\mathcal{H}$. To a codeword $\mathbf{w} = (k_1, k_2, \ldots, k_N)$ is associated the signal state $\mathbf{S}_\mathbf{w} = S_{k_1}\otimes S_{k_2}\cdots\otimes S_{k_N}$. A block code with M codewords is a mapping from a set of M messages $\{1,\ldots,M\}$ into a set of M codewords $\mathbf{w}_1,\ldots,\mathbf{w}_M$. A quantum decision scheme for such a code is a collection of M positive operators $\{\Pi_1,\Pi_2,\ldots,\Pi_M\}$ such that $\sum\Pi_i \leq \mathbb{1}$. The rate of the code is defined as $R = (\log M)/N$.

The probability that message m' is decoded when message m is transmitted is $P(m'|m) = \operatorname{Tr}\Pi_{m'}\mathbf{S}_{\mathbf{w}_m}$ and the total probability of error after sending message m is $P_{e,m} = 1 - \operatorname{Tr}(\Pi_m\mathbf{S}_{\mathbf{w}_m})$. We then define the maximum probability of error of the code $P_{e,max} = \max_m P_{e,m}$ and, for any positive R and integer N, we define $P^{(N)}_{e,max}(R)$ as the minimum maximum error probability over all codes of block length N and rate at least R.

For rates R smaller than the capacity of the channel, $P^{(N)}_{e,max}(R)$ goes to zero exponentially fast in N. The reliability function of the channel is defined as[1]

$$E(R) = \limsup_{N\to\infty} -\frac{1}{N}\log P^{(N)}_{e,max}(R). \quad (24)$$

The purpose of this section is to adapt the proof of the sphere packing bound in [7, Sec. IV] to the case of quantum channels. This results in the following theorem.

[1]It is known that the same function $E(R)$ results if in (24) one substitutes $P_{e,max}$ with the average probability of error over codewords $P_e = \sum_m P_{e,m}/M$, see for example [7], [3].

Theorem 2 (Sphere Packing Bound): For all positive rates R and all positive ε,

$$E(R) \leq E_{sp}(R-\varepsilon), \tag{25}$$

where $E_{sp}(R)$ is defined by the relations

$$E_{sp}(R) = \sup_{\rho \geq 0} \left[E_0(\rho) - \rho R\right] \tag{26}$$

$$E_0(\rho) = \max_{\mathbf{q}} E_0(\rho, \mathbf{q}) \tag{27}$$

$$E_0(\rho, \mathbf{q}) = -\log \operatorname{Tr} \left(\sum_{k=1}^{K} q_k S_k^{1/(1+\rho)} \right)^{1+\rho} \tag{28}$$

Remark 2: For some channels, the function $E_{sp}(R)$ can be infinite for R small enough. The role of the arbitrarily small constant ε is only important for one single value of the rate $R = R_\infty$, which is the infimum of the rates R such that $E_{sp}(R)$ is finite.

Proof: We follow closely the proof given in [7, Sec. IV] for the classical case. Some steps are clearly to be adapted to the quantum case and, since that proof is quite complicated, it would not be easy to explain how to do that without at least repeating the main steps of the proof. Hence, for the reader's convenience, we prefer to go through the whole proof used in [7] directly speaking in terms quantum channels and trying to simplify it as much as possible in view of the weaker results that we are pursuing with respect to [7, Th. 5] (we are here only interested in the asymptotic first order exponent, while in [7], bounds for fixed M and N are obtained).

The key point is using Fano's idea [2, Sec. 9.2] of bounding the probability of error for at least one codeword $\mathbf{w}_m$ by studying a binary hypothesis testing problem between $\mathbf{S}_{\mathbf{w}_m}$ and a dummy state $\mathbf{f}$, which is only used as a measure for the decision operator Π_m.

Here, we simplify the problem using the fact that for the study of $E(R)$ we can only consider the case of *constant composition* codes (see [2] [7]). This observation clearly holds also for classical-quantum channels, since it stems from the fact that the number of different compositions only grows polynomially in N, while the number of codewords grows exponentially. Hence, let c_k be the number of occurrences of symbol k in each word and define then q_k as the ratio c_k/N, so that the vector $\mathbf{q} = (q_1, q_2, \ldots, q_K)$ is obviously a probability distribution over the K input symbols.

Let now $\mathbf{f}$ be a state in $\mathcal{H}^{\otimes N}$. We will first apply Theorem 1 using one of the codewords as state ϱ and $\mathbf{f}$ as state ς. This will result in a trade-off between the rate of the code R and the probability of error $P_{e,max}$, where both quantities will be parameterized in the parameter s, a higher rate being allowed if a larger $P_{e,max}$ is tolerated and vice-versa. This trade-off depends of course on $\mathbf{q}$ and $\mathbf{f}$. We will later pick $\mathbf{f}$ properly so as to obtain the best possible bound for a given R valid for all compositions $\mathbf{q}$.

For any $m = 1 \ldots, M$, consider the binary hypothesis testing between $\mathbf{S}_{\mathbf{w}_m}$ and $\mathbf{f}$. We assume that $\mathbf{S}_{\mathbf{w}_m}$ and $\mathbf{f}$ have non-disjoint supports and define the quantity

$$\mu(s) = \log \operatorname{Tr} \mathbf{S}_{\mathbf{w}_m}^{1-s} \mathbf{f}^s. \tag{29}$$

Applying Theorem 1 with $\varrho = \mathbf{S}_{\mathbf{w}_m}$, $\varsigma = \mathbf{f}$ and $\Pi = \mathbb{1} - \Pi_m$, we find that for each s in $0 < s < 1$, either

$$\operatorname{Tr}\left[(\mathbb{1} - \Pi_m)\, \mathbf{S}_{\mathbf{w}_m}\right] > \frac{1}{8} \exp\left[\mu(s) - s\mu'(s) - s\sqrt{2\mu''(s)}\right] \tag{30}$$

or

$$\operatorname{Tr}\left[\Pi_m \mathbf{f}\right] > \frac{1}{8} \exp\left[\mu(s) + (1-s)\mu'(s) - (1-s)\sqrt{2\mu''(s)}\right]. \tag{31}$$

Note now that $\operatorname{Tr}\left[(\mathbb{1} - \Pi_m)\, \mathbf{S}_{\mathbf{w}_m}\right] = P_{e,m} \leq P_{e,max}$ for all m. Furthermore, since $\sum_{m=1}^{M} \Pi_m \leq \mathbb{1}$, for at least one value of m we have $\operatorname{Tr}\left[\Pi_m \mathbf{f}\right] \leq 1/M = e^{-NR}$. Choosing this particular m, we thus obtain from the above two equations that either

$$P_{e,max} > \frac{1}{8} \exp\left[\mu(s) - s\mu'(s) - s\sqrt{2\mu''(s)}\right] \tag{32}$$

or

$$R < -\frac{1}{N}\left(\mu(s) + (1-s)\mu'(s) - (1-s)\sqrt{2\mu''(s)} - \log 8\right) \tag{33}$$

In these equations we begin to see the aimed trade-off between the rate and the probability of error. It is implicit here in the definition of $\mu(s)$ that both equations depend on $\mathbf{S}_{\mathbf{w}_m}$ and $\mathbf{f}$. Since m has been fixed, we can drop its explicit indication and use simply $\mathbf{w}$ in place of $\mathbf{w}_m$ from this point on. We will now call $R(s, \mathbf{S}_{\mathbf{w}}, \mathbf{f})$ the right hand side of (33). This allows us to write $\mu'(s)$ in (32) in terms of $R(s, \mathbf{S}_{\mathbf{w}}, \mathbf{f})$ so that, taking the logarithm in equation (32), our conditions can be rewritten as either

$$R < R(s, \mathbf{S}_{\mathbf{w}}, \mathbf{f}) \tag{34}$$

or

$$\log \frac{1}{P_{e,max}} < -\frac{\mu(s)}{1-s} - \frac{sN}{1-s} R(s, \mathbf{S}_{\mathbf{w}}, \mathbf{f}) + 2s\sqrt{2\mu''(s)} + \frac{\log 8}{1-s}. \tag{35}$$

At this point, we exploit the fact that we are considering a fixed composition code. Since we want our result to depend only on the composition $\mathbf{q}$ and not on the particular sequence $\mathbf{w}$, we choose $\mathbf{f}$ so that the function $\mu(s)$ also only depends on the composition $\mathbf{q}$. We thus choose $\mathbf{f}$ to be the N-fold tensor power of a state f in $\mathcal{H}$, that is $\mathbf{f} = f^{\otimes N}$. With this choice, in fact, we easily check that, if $\mathbf{w}$ has composition $\mathbf{q}$,

$$\mu(s) = \log \operatorname{Tr} \mathbf{S}_{\mathbf{w}}^{1-s} \mathbf{f}^s \tag{36}$$

$$= N \sum_{k=1}^{K} q_k \log\left(\operatorname{Tr} S_k^{1-s} f^s\right). \tag{37}$$

Thus, $\mu(s)$ actually only depends on the composition $\mathbf{q}$ and on f, and not on the particular $\mathbf{w}$. It is useful to remember that since we assumed the supports of $\mathbf{f}$ and $\mathbf{S}_{\mathbf{w}}$ to be non-disjoint, the supports of S_k and f are not disjoint if $q_k > 0$, so that all terms in the sum are well defined. Setting

$$\mu_{k,f}(s) := \log\left(\operatorname{Tr} S_k^{1-s} f^s\right) \tag{38}$$

we thus have

$$\mu(s) = N \sum_k q_k \mu_{k,f}(s) \tag{39}$$

and hence, obviously,

$$\mu''(s) = N \sum q_k \mu''_{k,f}(s). \tag{40}$$

With the same procedure used to obtain (17) using the Nussbaum-Szkoła mapping (13), we see that for fixed s and f, $\mu''_{k,f}(s)$ is a variance of a finite random variable and it is thus a finite non-negative real number. Taking the largest of these numbers over k, say $C(s,f)$, we find that

$$\mu''(s) \leq NC(s,f). \tag{41}$$

We also observe that since $\mu''_{k,f}(s) \geq 0$ for all k, $\mu_{k,f}(s)$ is convex in s for all choices of f, a fact that will be useful later.

The essential point here is that the contribution of $\mu(s)$ and $\mu'(s)$ in our bounds will grow linearly in N, while the contribution of $\mu''(s)$ will only grow with $\sqrt{N}$. Hence, the terms involving $\mu''(s)$ become unimportant for large N. A formalization of this fact, however, is tricky. In [7] the effect of $\mu''(s)$ in the classical case is dealt with by bounding $s^2\mu''_{k,f}(s)$ by a constant uniformly over s and f, which allows the authors to proceed in deriving a bound on $P_{e,max}$ for all fixed M and N.

In our case, this procedure cannot be applied in a simple way (see [8] for details on the reasons) and we have to take at this point a slightly different approach, which will allow us to find a bound on $E(R)$ using the asymptotic regime $N \to \infty$. Simplifying again the notation in light of the previous observations, let us write $R(s,\mathbf{q},f)$ for $R(s,\mathbf{S_w},\mathbf{f})$. Using the obtained expression for $\mu(s)$, our conditions are either

$$R < R(s,\mathbf{q},f) \tag{42}$$

or

$$\frac{1}{N}\log\frac{1}{P_{e,max}} < -\frac{1}{1-s}\sum_k q_k\mu_{k,f}(s) - \frac{s}{1-s}R(s,\mathbf{q},f) + \frac{1}{N}\left(2s\sqrt{2\mu''(s)} + \frac{\log 8}{1-s}\right). \tag{43}$$

Now we come to the most critical step. Given a rate R, we want to bound $P_{e,max}$ for all codes. Here, we should choose s and f optimally depending on $\mathbf{q}$ and R, but we should then optimize the composition $\mathbf{q}$ in order to have a bound valid for all codes. This direct approach, even in the classical case, turns out to be very complicated (see [2, Sec. 9.3 and 9.4, pag. 188-303] for a detailed and however instructive analysis). The authors in [7] thus proceed in a more synthetic way by stating the resulting optimal f and $\mathbf{q}$ as a function of s and then proving that this choice leads to the desired bound. Here, we will follow this approach showing that the same reasoning can be applied also to the case of quantum channels. It is important to point out that it is not possible to simply convert the quantum problem to the classical one using the Nussbaum-Szkoła mapping (13) directly on the states S_k and f and then using the construction of [7, eqs. (4.18)-(4.20)] on the obtained classical distributions. In fact, in (13), even if one of the two states is kept fixed and only the other one varies, *both* distributions vary. Thus, even if f is kept fixed, the effect of varying S_k for the different values of k would not be compatible with the fact that in [7, eq. (4.20)] a fixed $\mathbf{f}_s$ (in that notation) is defined, which is not supposed to depend on k. Fortunately, it is instead possible to exactly replicate the steps used in [7] by correctly reinterpreting the construction of f and $\mathbf{q}$ in the quantum setting.

For a fixed s in the interval $0 < s < 1$, consider the quantity

$$E_0\left(\frac{s}{1-s},\mathbf{q}\right) = -\log\mathrm{Tr}\left(\sum_k q_k S_k^{1-s}\right)^{1/(1-s)} \tag{44}$$

and call $\mathbf{q}_s = (q_{1,s},\ldots,q_{K,s})$ the choice of $\mathbf{q}$ that maximizes this expression. As observed by Holevo[2] [4, eq. (38)], $\mathbf{q}_s$ satisfies the conditions

$$\mathrm{Tr}\left(S_k^{1-s}\alpha_s^{s/(1-s)}\right) \geq \mathrm{Tr}\left(\alpha_s^{1/(1-s)}\right); \qquad k = 1,\ldots,K \tag{45}$$

where

$$\alpha_s = \sum_{k=1}^{K} q_{k,s} S_k^{1-s}. \tag{46}$$

Furthermore, equation (45) is satisfied with equality for those k with $q_{k,s} > 0$, as can be verified by multiplying it by $q_{k,s}$ and summing over k.

Define now

$$f_s = \frac{\alpha_s^{1/(1-s)}}{\mathrm{Tr}\,\alpha_s^{1/(1-s)}}. \tag{47}$$

Since we can choose s and f freely, we will now tie the operator f to the choice of s, using f_s for f. We only have to keep in mind that $\mu'(s)$ and $\mu''(s)$ are computed by holding f fixed. Note further that we fullfill the requirement that f and S_k have non-disjoint supports, since the left hand side in (45) must be positive for all k.

As in [7, eqs (4.21)-(4.22)], we see that, using f_s in place of f in the definition of $\mu_{k,f}(s)$, we get

$$\mu_{k,f_s}(s) = \log\mathrm{Tr}\left(S_k^{1-s}\alpha_s^{s/(1-s)}\right) - s\log\mathrm{Tr}\,\alpha_s^{1/(1-s)}. \tag{48}$$

Using (45) we then see that

$$\begin{aligned}\mu_{k,f_s}(s) &\geq (1-s)\log\mathrm{Tr}\,\alpha_s^{1/(1-s)} &(49)\\ &= -(1-s)E_0\left(\frac{s}{1-s},\mathbf{q}_s\right) &(50)\\ &= -(1-s)E_0\left(\frac{s}{1-s}\right) &(51)\end{aligned}$$

with equality if $q_{k,s} > 0$. Here, we have used the definitions (46), (28) and (27), and the the fact that $\mathbf{q}_s$ maximizes (44). Thus, with the choice of $f = f_s$, equations (42) and (43) can be rewritten as (for each s) either

$$R < R(s,\mathbf{q},f_s) \tag{52}$$

or

$$\frac{1}{N}\log\frac{1}{P_{e,max}} < E_0\left(\frac{s}{1-s}\right) - \frac{s}{1-s}R(s,\mathbf{q},f_s) + \frac{2s\sqrt{2}}{\sqrt{N}}\sqrt{\sum_k q_k\mu''_{k,f_s}(s)} + \frac{\log 8}{(1-s)N} \tag{53}$$

[2]The variable s in [4] corresponds to our $s/(1-s)$, that we call ρ here in accordance with the consolidated classical notation.

where

$$R(s,\mathbf{q},f_s) = -\sum_k q_k \mu_{k,f_s}(s) - (1-s)\sum_k q_k \mu'_{k,f_s}(s) + \frac{1}{\sqrt{N}}(1-s)\sqrt{2\sum_k q_k \mu''_{k,f_s}(s)} + \frac{1}{N}\log 8. \quad (54)$$

Using the same procedure used in [7, pag. 100-102], invoking the strict convexity of $\mathrm{Tr}(\alpha^{1/(1-s)})$ in α for $0 < s < 1$, it can be proved that $R(s,\mathbf{q},f_s)$ is a continuous function of s. Thus, for fixed R, we can only have three possibilities:

1) $R = R(s,\mathbf{q},f_s)$ for some s in $(0,1)$;
2) $R > R(s,\mathbf{q},f_s)$ $\forall s \in (0,1)$;
3) $R < R(s,\mathbf{q},f_s)$ $\forall s \in (0,1)$.

Our conditions are slightly different from those in [7] due to the fact that we have not been able to bound uniformly the second derivatives $\mu''_{k,f_s}(s)$ for $s \in (0,1)$. For this same reason, dealing with these possibilities for a fixed code is more complicated in our case than in [7]. Thus, we have to depart slightly from [7]. Due to space limitation, we can give here only a concise explanation that should be sufficient when integrated with [7], the interested reader can find more precise technical details in [8].

Instead of considering a fixed code of block length N, consider sequences of codes. From the definition of $E(R)$ in (24), it is obvious that there exists a sequence of codes of block-lengths $N_1, N_2, \ldots, N_n, \ldots,$ and rates $R_1, R_2, \ldots, R_n, \ldots$ such that $R = \lim_n R_n$ and

$$E(R) = \lim_{n\to\infty} -\frac{1}{N_n}\log P_{e,max}^{(N_n)}(R). \quad (55)$$

Each code of the sequence will in general have a different composition $\mathbf{q}_n$ but must anyway fall in one of the above three cases. Thus, one of those cases is verified infinitely often. Since the compositions $\mathbf{q}_n$ are in a bounded set, there exists a subsequence of codes such that $\mathbf{q}_n$ converge to, say, $\bar{\mathbf{q}}$. Thus, we can directly assume this subsequence is our own sequence and safely assume that $\mathbf{q}_n \to \bar{\mathbf{q}}$.

Suppose now that case (1) is verified infinitely often. Thus, for infinitely many n, there is an $s = s_n$ in the interval $0 < s < 1$ such that $R_n = R(s,\mathbf{q}_n,f_{s_n})$. Hence, since the values s_n are in the interval $(0,1)$, there must exists an accumulation point for the s_n in the closed interval $[0,1]$. We will first assume that an accumulation point $\bar{s}$ exists satisfying $0 < \bar{s} < 1$. A subsequence of codes then exists with the s_n tending to $\bar{s}$. Let this subsequence be our new sequence. We can first substitute $R(s_n,\mathbf{q}_n,f_{s_n})$ with R_n in (53). Letting then $n \to \infty$, we find that $R_n \to R$ and the last two terms on the right hand side of (53) vanish. Hence, we obtain

$$\begin{aligned} E(R) &\le E_0\left(\frac{\bar{s}}{1-\bar{s}}\right) - \frac{\bar{s}}{1-\bar{s}}R & (56)\\ &\le \sup_{\rho\ge 0}\left(E_0(\rho) - \rho R\right) & (57)\\ &= E_{sp}(R). & (58)\end{aligned}$$

Suppose now that either case (2) above is verified infinitely often, or that case (1) is with the only accumulating point $\bar{s} = 0$ for the values s_n. Given any $\varepsilon_1 > 0$, for any fixed $s \in [\varepsilon_1, 1)$ we must have $R(s,\mathbf{q}_n,f_s) \le R_n$ infinitely often. Since condition (52) is not satisfied, (53) must be satisfied infinitely often for any fixed $s \in [\varepsilon_1, 1)$. Making $n \to \infty$ we can get rid again of the last two terms in (53) and have

$$E(R) \le E_0\left(\frac{s}{1-s}\right) - \frac{s}{1-s}R(s,\bar{\mathbf{q}},f) \quad (59)$$

Letting then $\varepsilon_1 \to 0$, we can let $s \to 0$ as well and find that $E(R) \le 0$. Thus, surely $E(R) \le E_{sp}(R)$ proving the theorem in this case (see [8] for more details here).

Suppose finally that either case (3) above is verified infinitely often, or that case (1) is with the only accumulating point $\bar{s} = 1$ for the values s_n. Given any $\varepsilon_1 > 0$, for all $s \in (0, 1-\varepsilon_1]$, the inequality $R_n < R(s,\mathbf{q}_n,f_s)$ is verified infinitely often, so that we can take this time the limit $n \to \infty$ in (52) and (54). Proceeding exactly as in [7], we can then use the convexity of $\mu_{k,f}(s)$ and (51) to prove that

$$R \le \frac{1-s}{s}E_0\left(\frac{s}{1-s}\right) \quad (60)$$

for all $s \in (0, 1-\varepsilon_1]$. Setting $\rho = s/(1-s)$ and $\rho_1 = (1-\varepsilon_1)/\varepsilon_1$, this implies that for any $\varepsilon_2 > 0$,

$$\begin{aligned} E_{sp}(R-\varepsilon_2) &= \sup_{\rho\ge 0}\left(E_0(\rho) - \rho(R-\varepsilon_2)\right)\\ &\ge \sup_{0\le\rho\le(1-\varepsilon_1)/\varepsilon_1}\left(E_0(\rho) - \rho(R-\varepsilon_2)\right)\\ &\ge \sup_{0\le\rho\le(1-\varepsilon_1)/\varepsilon_1}\rho\varepsilon_2\\ &= \frac{(1-\varepsilon_1)\varepsilon_2}{\varepsilon_1}.\end{aligned}$$

which is arbitrarily large for any ε_2 if ε_1 is small enough. This proves that $E_{sp}(R-\varepsilon_2)$ is unbounded for arbitrarily small ε_2 and thus surely $E(R) \le E_{sp}(R-\varepsilon_2)$, concluding the proof. ■

REFERENCES

[1] M. V. Burnashev and A. S. Holevo, "On reliability function of quantum communication channel," *Probl. Peredachi Inform.*, vol. 34, no. 2, pp. 1–13, 1998.

[2] R. Fano, *Transmission of Information: A Statistical Theory of Communication*. Wiley, New York, 1961.

[3] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, vol. IT-11, no. 1, pp. 3–18, January 1965.

[4] A. S. Holevo, "Reliability function of general classical-quantum channel," *IEEE Trans. Inform. Theory*, vol. IT-46, no. 6, pp. 2256 –2261, 2000.

[5] M. Nussbaum and A. Szkoła, "The chernoff lower bound for symmetric quantum hypothesis testing," *Ann. Statist.*, vol. 37, no. 2, pp. 1040–1057, 2009.

[6] K. M. R. Audenaert, M. Nussbaum, A. Szkoła, and F. Verstraete, "Asymptotic error rates in quantum hypothesis testing," *Comm. in Math. Phys.*, vol. 279, pp. 251–283, 2008.

[7] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding in discrete memoryless channels. I," *Information and Control*, vol. 10, pp. 65–103, 1967.

[8] M. Dalai, "Sphere packing and zero-rate bounds to the reliability of classical-quantum channels," *arXiv:1201.5411v1*.

[9] C. W. Helstrom, *Quantum detection and estimation theory*. Academic Press, New York, 1976.

[10] A. S. Holevo, "An analog of the theory of statistical decisions in noncommutative theory of probability,," *Tr. Mosk. Matemat. Obshchest*, vol. 26, pp. 133–149, 1972.